Chapter 2

Describing Distributions with Numbers

Intended Learning Outcomes

- Compute and interpret **measures of central tendency**: mean, median, and mode
- Explain the difference between **resistant** and **non-resistant** measures
- Compute and interpret **measures of variability**: IQR, variance, and standard deviation
- Construct and interpret standard and **modified boxplots**
- Identify potential **outliers** using the $1.5 \times$ IQR rule
- Choose appropriate **summary statistics** based on distribution shape

Why Numerical Summaries Matter

Numerical Summary

A **numerical summary** reduces an entire distribution to a few key numbers that capture its essential features: where the data is **centred** and how much it **varies**.

Two Fundamental Questions:

1. Where is the distribution **centred**? (Central tendency)
2. How **spread out** is the distribution? (Variability)

Indices and Summation

of rows or observations

n ← Stop at index n

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \cdots + x_n$$

Add this variable

$i = 1$ ← Start at index 1

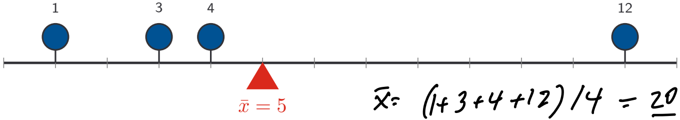
Key Point: The symbol Σ (capital Greek sigma) is an instruction to **add up** everything that follows it.

Measuring Centre: The Mean

Mean (Average)

The **mean** of n observations $x_1, x_2, \ldots, x_n$ is the sum of all values divided by the number of observations.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$



$\bar{x} = (1+3+4+12)/4 = \frac{20}{4} = 5$

Figure: The mean as a balance point. Data: {1, 3, 4, 12}

Example 2.1: Calculating a Sum

Find $\sum_{i=1}^{3} x_i$ for the $\underbrace{\texttt{countries\_visited}}_{X}$ variable.

ID	x
1	2
2	5
3	1

$$\sum_{i=1}^{3} X_i = X_1 + X_2 + X_3$$
$$= 2 + 5 + 1 = 8$$

Example 2.2: Calculating a Sum and Mean

Five students reported the number of books they read last year.

(a) Find $\sum_{i=1}^{5} x_i$

(b) Calculate the mean $\bar{x}$

Student	x
1	3
2	7
3	2
4	10
5	5

$n=5$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3+7+2+10+5}{5} = \frac{27}{5} \doteq 5.4$$

5.4 books is the average # of books read

Robustness of a Measure

Resistant Measure

A statistical measure is **resistant** (or robust) if it is not sensitive to the influence of a few extreme values (outliers). A resistant measure focuses on the center of the data rather than the tails.

Why does this matter?

- Real data sometimes contains unusual values (typos, measurement errors, genuinely extreme cases)
- Choosing the wrong summary can give a misleading picture

Example: The Mean is Non-Resistant

Ages of five book club members:

```
22, 25, 28, 30, 32
```

(a) Calculate the mean age.

(b) A 95-year-old joins. Calculate the new mean.

(c) Does the new mean represent "typical"?

a) $\bar{x} = 27.4$ (Check)

b) $\frac{(22+25+28+30+32+95)}{6} = 38.7 \leftarrow$ new mean

c) The average is no longer representative of the typical book club member.

Measuring Centre: The Mode

Mode

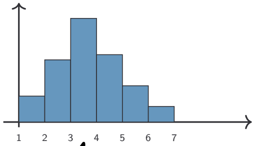
The **mode** is the value that appears the most in a dataset. It is the only measure of centre that we **can** use for categorical data.

Types of Modality:

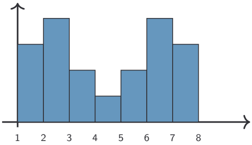
- Unimodal: One peak
- Bimodal: Two peaks
- Multimodal: More than two peaks
- No mode: No value repeats

Visualizing Modality

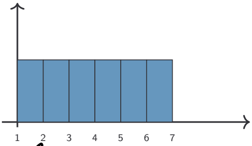
In each histogram, identify the modes:



The mode is [3,4)



Bimodal: [2,3) and [6,7)



No modes

Example 2.7: Finding the Mode

For each dataset, identify the mode(s):

Dataset	Mode(s)
$\{4, 7, 7, 7, 9, 12\}$	7
$\{2, 4, 6, 8, 10\}$	N/A (all observations occur same # of times)
$\{1, 3, 3, 4, 5, 6, 6, 9\}$	3, 6
T-shirt sizes: {S, M, L, M, L, XL, L}	L

Measuring Centre: The Median

Median

The **median** is the middle value when data are arranged in order. It divides the distribution so that half of observations fall below and half fall above.

Finding the Median

1. Sort the data from smallest to largest.
2. If n is odd: The median is the middle value (position $\frac{n+1}{2}$).
3. If n is even: The median is the average of the two middle values.

Example 2.9: Finding the Median (Odd n)

Find the median of:

```
5, 21, 4, 8, 22, 11, 5
```

3 obs. 3 obs.

Sort:

$n=7$

4	5	5	8	11	21	22
	5	5	8	11	21	
		5	8	11		
			8			

3 observations 3 observations

Example 2.10: Finding the Median (Even n)

Find the median of:

$$5,\ 6,\ 7,\ 7.5,\ 7.5,\ 8,\ 8,\ 8$$

$\underbrace{5 \quad 6 \quad 7 \quad 7.5}_{4 \text{ obs.}}$ | $\underbrace{7.5 \quad 8 \quad 8 \quad 8}_{4 \text{ obs.}}$ (7.5)

6 7 7.5 7.5 8 8

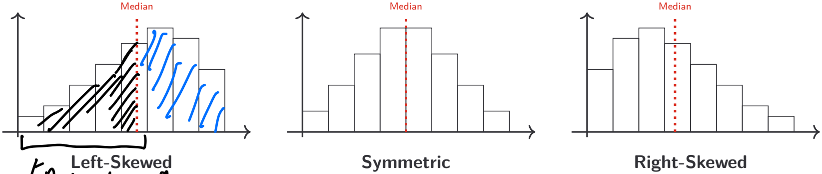
7.5 7.5

$$\text{Median} = \frac{7.5 + 7.5}{2} = 7.5$$

When n is even, we average the two values in the middle.

Visualizing the Median on Histograms

50% of dist^n

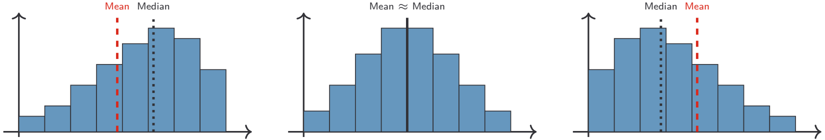


50% of dist^n

Key Point: The median divides the area under the histogram in half:
50% of data are to the left, 50% to the right.

Mean vs. Median and Skewness

Note: This is not always true, but we will use it as a rule of thumb as it is true in many dist^ns one would encounter.



Left-Skewed
Mean < Median

Symmetric

Right-Skewed
Mean > Median

If left skewed, then mean < median
right skewed, then mean > median
symmetric, then mean ≈ median

From Centre to Spread: The Missing Piece

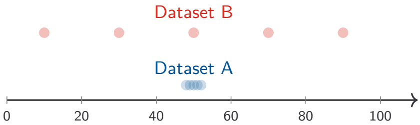
Motivating Question: Consider two datasets with the same mean of 50:

Dataset A:

`48, 49, 50, 51, 52`

Dataset B:

`10, 30, 50, 70, 90`



Key Point: The mean alone cannot distinguish these very different distributions. We need measures of **spread** to complete the picture.

Quartiles

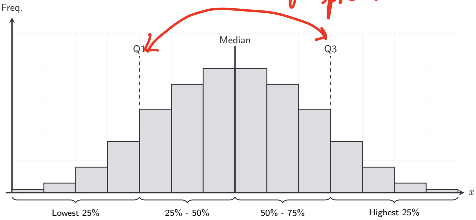
Quartiles

Quartiles divide a sorted dataset into four equal parts.

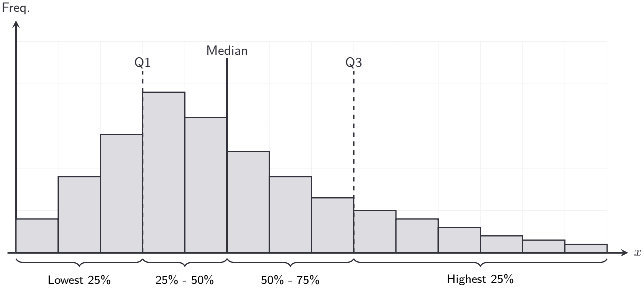
- **First Quartile (Q_1)**: 25% of data fall below this value
- **Second Quartile (Q_2)**: The median (50% below, 50% above)
- **Third Quartile (Q_3)**: 75% of data fall below this value

Quartiles help describe the spread and center of a distribution.

Notion of spread



Quartiles



Quartiles and the Interquartile Range

Calculating Quartiles

1. **Sort** the data from smallest to largest.
2. Find the **Median (Q_2)**.
3. **If n is odd:** Exclude the median value.
   - Lower Half: All values strictly below the median.
   - Upper Half: All values strictly above the median.
4. **If n is even:** Split the data into two equal halves.
5. **Q_1:** The median of the Lower Half.
6. **Q_3:** The median of the Upper Half.

Example 2.13: Computing Quartiles (Odd n)

Find Q_1, Q_2 (median), and Q_3 for:

```
3, 5, 8, 12, 15, 28, 35
```

Q_2 = median
= 12

$\underbrace{3 \quad 5 \quad 8}$ 12 $\underbrace{15 \quad 28 \quad 35}$

Q_1 = median of the lower half
= 5

Find median of
3 5 8

Q_3 = median of the upper half

15 28 35

Q_3 = 28

Example 2.14: Computing Quartiles (Even n)

Find Q_1, Q_2 (median), and Q_3 for:

$$\underbrace{2,\ 4,\ 7,\ 9}_{LH},\ \overset{9.5}{|}\ \underbrace{10,\ 12,\ 15,\ 18}_{UH}$$

$Q_2 = \frac{9+10}{2} = 9.5$

Q_1 = median of LH
= median of 2, 4, 7, 9
$= \frac{4+7}{2} = 5.5$

Q_3 = median of UH
= median of 10, 12, 15, 18
$= \frac{12+15}{2} = 13.5$

$Q_1 = 5.5$, $Q_2 = 9.5$, $Q_3 = 13.5$

Last time

- Goal: • summarize dataset using 2 numbers
  - center
  - spread

sum of all x_i

- Center: • mean or average •: $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
- Mode: most frequently occurring value
- Median: middle value
- Special case of Quartiles

$$\text{Median} = Q_2.$$

The Five-Number Summary

Five-Number Summary

The **five-number summary** of a variable consists of the following five values, arranged in order from smallest to largest:

$$\text{Min} \qquad Q_1 \qquad \text{Median} \qquad Q_3 \qquad \text{Max}$$

Example: Coffee expenses ($)

```
3, 5, 8, 12, 15, 28, 35
```

Find the five-number summary.

LH $\qquad Q_2 = 12 \qquad$ UH

$\underbrace{3 \quad 5 \quad 8}_{}\quad 12 \quad \underbrace{15 \quad 28 \quad 35}_{}$

$Q_1 = 5 \qquad\qquad Q_3 = 28$

$\text{Min} = 3, \qquad \text{Max} = 35$

Example 2.15: The Interquartile Range

Interquartile Range (IQR)

The **interquartile range (IQR)** is the range of the middle 50% of the data:

$$\text{IQR} = Q_3 - Q_1$$

Example: Coffee expenses ($)

3, 5, 8, 12, 15, 28, 35

Compute the IQR.

Recall: $Q_1 = 5$, $Q_3 = 28$

$\text{IQR} = 28 - 5$

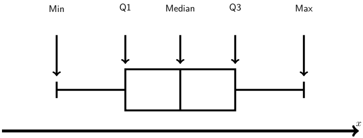
$= 23$

Boxplot

Boxplot

A **boxplot** is a graphical summary of a dataset based on the five-number summary:

- The **box** spans from the first quartile (Q_1) to the third quartile (Q_3).
- A line inside the box marks the **median**.
- **Whiskers** extend from the box to the minimum and maximum values.



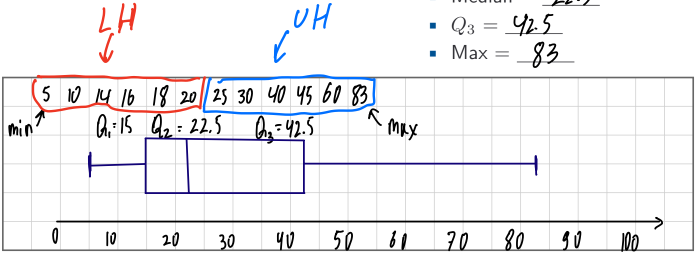
Example 2.18: Drawing a Boxplot

Data:

```
5, 10, 14, 16, 18, 20, 25, 30, 40,
45, 60, 83
```

Five-Number Summary:

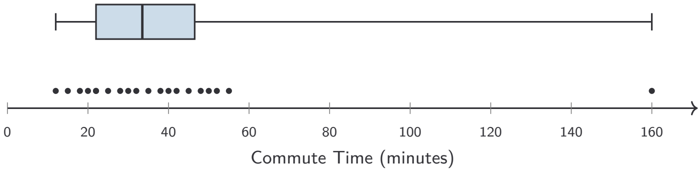
- Min = 5
- Q_1 = 15
- Median = 22.5
- Q_3 = 42.5
- Max = 83



Example 2.21: The Challenge with Boxplots

Data (minutes):

```
12, 15, 18, 20, 22, 22, 25, 28, 30, 32, 35, 38, 40, 42, 45, 48, 50, 52, 55,
160
```



Identifying Outliers: The $1.5 \times$ IQR Rule

Potential Outliers

A value is a **potential outlier** if it falls more than 1.5 IQR beyond the quartiles:

- **Lower Fence:** $Q_1 - 1.5 \times \text{IQR}$
- **Upper Fence:** $Q_3 + 1.5 \times \text{IQR}$

Any value lower the lower fence or higher the upper fence is flagged as an outlier.

This choice '1.5' IQR is arbitrary and was popularized by John Tukey in the 1970s. Other methods for identifying outliers exist.

Modified Boxplot

Modified Boxplot

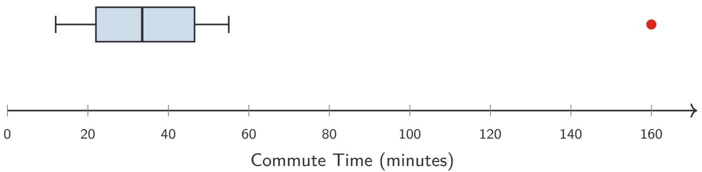
A **modified boxplot**

- displays potential outliers as ______________ and
- its whiskers extend only to the most extreme values *within* the fences.

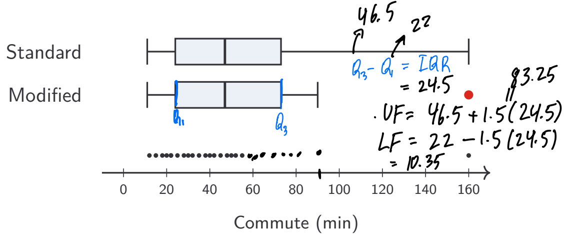
Example 2.21: Modified Boxplot (Daily Commute Times)

Data (minutes):

```
12, 15, 18, 20, 22, 22, 25, 28, 30, 32, 35, 38, 40, 42, 45, 48, 50, 52, 55,
160
```



Standard vs. Modified Boxplot



Example 2.22: Modified Boxplot with Two Outliers

Draw a modified boxplot for the following data:

```
5, 45, 48, 50, 52, 55, 58, 60, 62, 110
```

Upper whisker

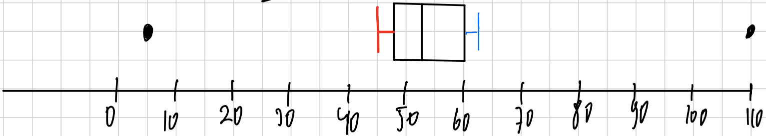
$$IQR \Rightarrow 1.5\,IQR = 18$$

$$UF = Q_3 + 1.5\,IQR = 60 + 18 = 78$$

$$LF = Q_1 - 1.5\,IQR = 48 - 18 = 30$$

$$Q_1 = 48,\quad Q_3 = 60,\quad Q_3 - Q_1 = 12 =$$

$$Q_2 = \frac{52+55}{2} = \frac{107}{2} = 53.5$$



Example 2.22: Modified Boxplot with Two Outliers

Draw a modified boxplot for the following data:

Upper whisker

```
5, 45, 48, 50, 52, 55, 58, 60, 62, 110
```

IQR ⇒ 1.5 IQR = 18

UF = $Q_3 + 1.5\,IQR$

45, 48, 50, 52, 55, 58, 60, 62

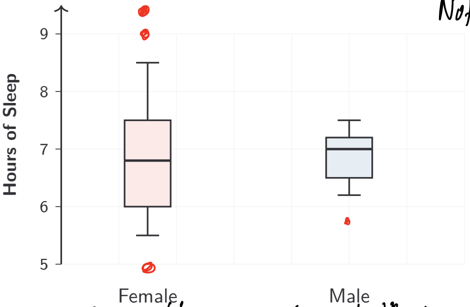
$Q_1 = 49$

$Q_3 = 59$

Example 2.23: Comparing Distributions with Boxplots

Who sleeps more?

(Modified)



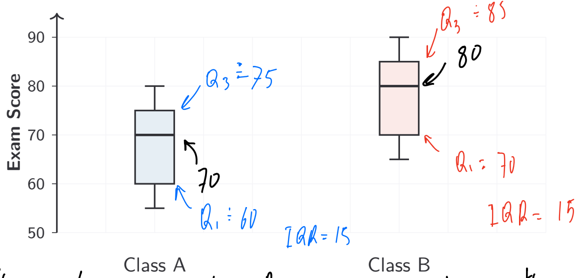
Note:
dist^n = distribution
f^n = function

How do these compare? Outliers: neither dist^n has outliers
- Center: the typical male sleeps more than the typical female
- Spread: men have a more consistent (less variable) sleep dist^n.

Example 2.24: Comparing Exam Scores: Two Classes

Same spread, different centres

Context: Final exam scores (out of 100) for two statistics classes.



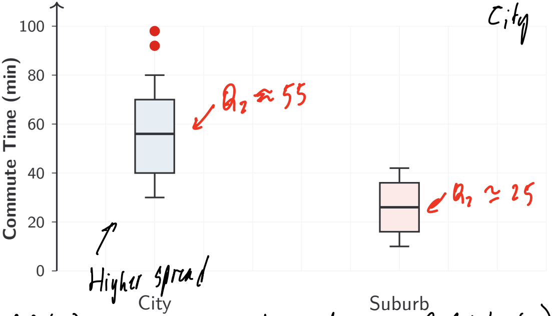
- Center: the median in class B is 10 marks higher than in A
- Spread: the spread is approximately the same: 15
- Outliers: no outliers in either plot.

Example 2.25: Comparing Commute Times: City vs. Suburb

Different centres, different spreads, with outliers

Context: Daily commute times (minutes) for workers in two locations.

Outliers:
City has two outliers, one at 92 and one at 99
No outliers for suburbs



$Q_2 \approx 55$

$Q_2 \approx 25$

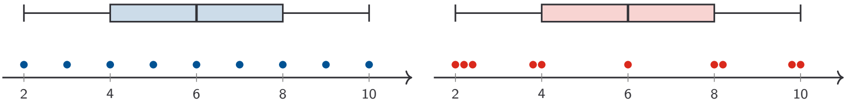
Higher spread

Center: Q_2(City) is 30 mis higher than Q_2(Suburbs)

Spread: IQR(City) = 70 - 40 = 30 , IQR(suburbs) = 35 - 18 = 17

Limitation: Same Boxplot, Different Distributions

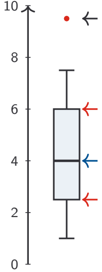
Boxplots hide details about distribution ~~shape~~ behaviour



Key Point: Two very different distributions can produce identical boxplots

Reading a Boxplot: Centre, Spread, and Outliers

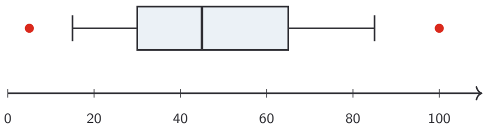
A boxplot reveals all three key features at once



Key Point: A single boxplot shows where the data is centred, how spread out it is, and identifies potential outliers.

Example 2.26: Practice Reading a Boxplot

Daily study time (minutes) for 50 students



(a) What is the median study time?

____

(b) What is the IQR?

____

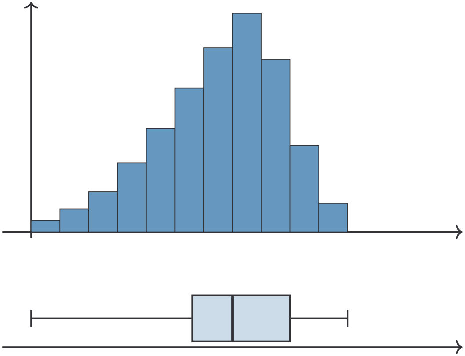
(c) Identify the outliers.

____

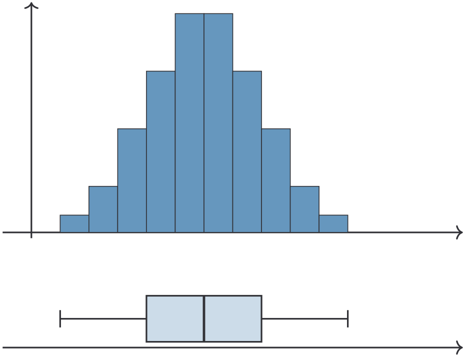
(d) Is the distribution symmetric?

____

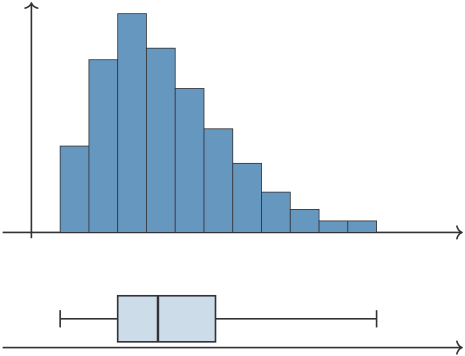
Distribution Shape: Left Skewed



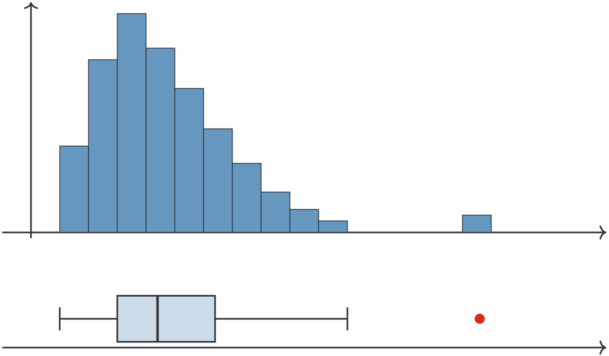
Distribution Shape: Symmetric



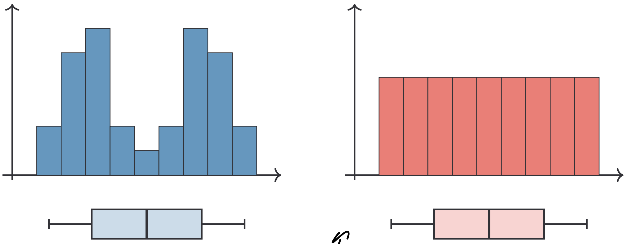
Distribution Shape: Right Skewed



Distribution Shape: Right Skewed with Outlier



Limitation: Boxplots



Different dist^n's with the same boxplot.

Beyond Quartiles: A More Precise Measure of Spread

The IQR tells us about the middle 50% of data.

Limitation of IQR:

- Ignores 50% of observations

We could, alternatively

- Measure how far each observation is from the centre
- Average these distances
- This results in the **standard deviation**

Key Point: The standard deviation uses *all* data points to quantify spread, making it a more comprehensive (but less resistant) measure.

Measuring Variability: Deviation

Deviation

A **deviation** is the difference between an observation and the mean

$$\text{deviation} = x_i - \bar{x}$$

mean

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Example 2.28: Visualizing Deviations from the Mean

Data:

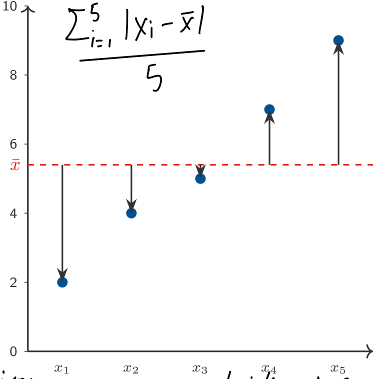
```
2, 4, 5, 7, 9
```

Mean: $\bar{x} = 5.4$

x_i	Deviation	Deviation^2
2	$2 - 5.4 = -3.4$	$(-3.4)^2$
4	$4 - 5.4 = -1.4$	$(-1.4)^2$
5	$5 - 5.9 = -0.4$	$(-0.4)^2$
7	$7 - 5.4 = 1.6$	$(1.6)^2$
9	$9 - 5.4 = 3.6$	$(3.6)^2$

Average of the |deviations|:

$$\frac{\sum_{i=1}^{5} |x_i - \bar{x}|}{5}$$



Problem: $\sum_{i=1}^{5} (x_i - \bar{X}) = 0$

Total of the deviations $\Rightarrow$ average deviation is 0

Measuring Variability: Variance

Variance

(squared deviation)

The **variance** measures the average deviation^2 from the mean

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

We divide by $n-1$ (not n) to get an unbiased estimate of the population variance.

Example 2.30: Computing Variance

Data:

```
2, 4, 5, 7, 9
```

Mean: $\bar{x} = 5.4$

Find the variance.

Example 2.31: Variance Practice

Find the variance for the following dataset

Variance $= s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

```
3, 6, 7, 8, 13
```

$n = 5$

$2 - 5.4 = -3.4$
$4 - 5.4 = -1.4$
$5 - 5.4 = -0.4$
$7 - 5.4 = 1.6$
$9 - 5.4 = 3.6$

Deviation^2
$(-3.4)^2 = 11.56$
$(-1.4)^2 = 1.96$
$(-0.4)^2 = 0.16$
$(1.6)^2 = 2.56$
$(3.6)^2 = 12.96$

$s^2 = \frac{1}{4}(11.56 + 1.96 + 0.16 + 2.56 + 12.96)$

$= \frac{1}{4}(29.2) = 7.3$

Variance

Standard Deviation

The **standard deviation** is the square root of the variance:

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

It has the same units as the original data, making it more interpretable than variance.

E.g. if x_i measure height in cm, then the std dev. is in cm.
the variance is in cm^2

Example 2.33: Computing Standard Deviation

Find the standard deviation for the following data

2, 4, 5, 7, 9

	Deviation^2
$2-5.4 = -3.4$	$(-3.4)^2 = 11.56$
$4-5.4 = -1.4$	$(-1.4)^2 = 1.96$
$5-5.4 = -0.4$	$(-0.4)^2 = 0.16$
$7-5.4 = 1.6$	$(1.6)^2 = 2.56$
$9-5.4 = 3.6$	$(3.6)^2 = 12.96$

$$s^2 = \frac{1}{4}(11.56 + 1.96 + 0.16 + 2.56 + 12.96)$$

$$= \frac{1}{4}(29.2) = 7.3$$

Variance

$$\sqrt{s^2} = s = \sqrt{7.3} \doteq 2.7$$

$\therefore$ values are typically 2.7 units from the mean

Example 2.34: Same Mean, Different Spread

Two sections of a statistics course have quiz scores:

Section A:

```
73, 74, 75, 76, 77
```

Mean = 75

$s \approx$ 1.58

check

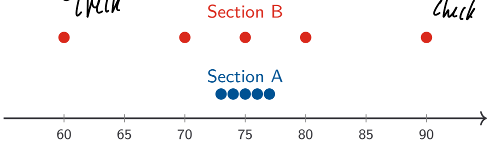
Section B:

```
60, 70, 75, 80, 90
```

Mean = 75

$s \approx$ 11.18

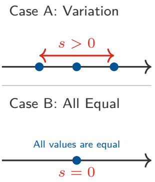
check

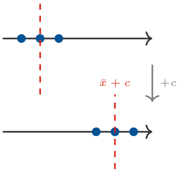


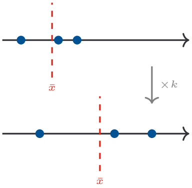
Properties of Standard Deviation

Key Point: The standard deviation satisfies the following properties:

- $s \geq 0$; $s = 0$ only if all values are equal
- Adding/subtracting a constant does not change s
- Multiplying all values by k changes s to $|k|s$







Example 2.35: Adding a Constant to All Values

Original scores:

```
65, 70, 75, 80, 85
```

Mean = 75
$s \approx 7.91$

After adding 10 points:

```
75, 80, 85, 90, 95
```

Mean = 85
$s \approx$ 7.91

Center summaries

- Mean
- Median
- ~~Mode~~

Spread summaries:

- IQR
- Std dev (variance)

Question: When do we use which?

Context dependent: If there are outliers, use median and IQR
If the distn is skewed. use median and IQR
Otherwise, use the mean and std dev.

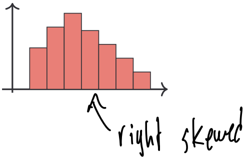
Choosing Measures of Centre and Variability

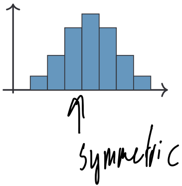
Median & IQR

Use when distribution is **skewed** or when **outliers** are present.

Mean & Standard Deviation

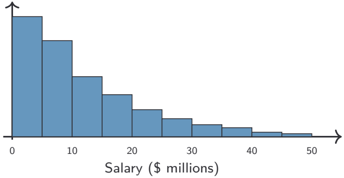
Use when distribution is **symmetric** without outliers.





Example 2.36: NBA Player Salaries (2024–25)

The histogram below shows the distribution of NBA player salaries for the 2024–25 season ($n = 450$ players).



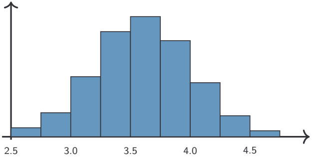
Summary Statistics:
Mean: $10.2M
Median: $4.5M
SD: $11.8M
IQR: $12.1M

(a) Is this distribution symmetric or skewed?
right skewed

(b) Which measures should we report?
Centre: median
Spread: IQR

Example 2.37: Boston Marathon Finishing Times (2024)

Context: Finishing times (hours) for 500 randomly sampled runners from the 2024 Boston Marathon.



Summary Statistics:
Mean: 3.72 hrs
Median: 3.68 hrs
SD: 0.48 hrs
IQR: 0.65 hrs

(a) Is this distribution symmetric or skewed?

Symmetric

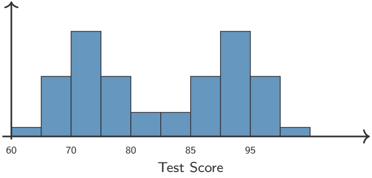
(b) Which measures should we report?

Centre: Mean

Spread: Std dev.

Example 2.38: Bimodal Symmetric Distribution

A combined sample of test scores (out of 100) from two sections of a course is shown below:



Summary Statistics:
Mean: 80.0
Median: 80.0
Mode: 72 and 90 (bimodal)
SD: 9.4

(a) Describe the shape.

Symmetric

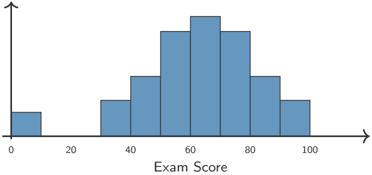
(b) What might explain this pattern?

(c) Which statistics are appropriate?

mean, std. dev.

Example 2.39: Symmetric Distribution with Outliers

A class of students took an exam, and their scores (out of 100) are shown below:



Summary Statistics:
Mean: 81.0
Median: 86.0
SD: 19.8
IQR: 16.0

(a) Describe the shape of the distribution.

left skewed or symmetric

(b) Identify the outlier.

____

(c) Which statistics should we report?

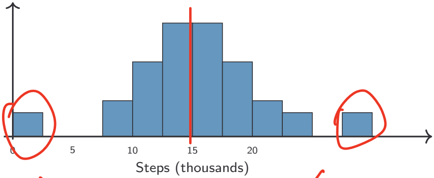
Centre: Median

Spread: IQR

Example 2.40: Symmetric Distribution with Outliers on Both Sides

Daily step counts with extreme values

A fitness tracker recorded daily step counts for a group of individuals over a month.



Outliers

Summary Statistics:

Mean: ~~10,913~~ 15000

Median: 10,800

SD: 3,445

IQR: 2,600

(a) Describe the shape of the distribution.

Symmetric

(b) Identify the outliers.

(c) Which statistics should we report?

Centre: Median

Spread: IQR

Example 2.41: Toronto Home Prices (December 2025)

Prices ($ thousands) for 15 homes sold in a Toronto neighbourhood.

```
685, 720, 745, 780, 795, 810, 825, 850, 875, 890, 920, 985, 1050, 1180,
2450
```

(a) Calculate the five-number summary.

(b) Check for outliers using the $1.5 \times \text{IQR}$ rule.

(c) Which statistics best describe "typical" prices?

Example 2.42: AP Statistics Exam Scores (2024)

A sample of 12 students' AP Statistics exam scores (out of 100):

```
72, 75, 78, 79, 81, 82, 84, 85, 87, 88, 90, 91
```

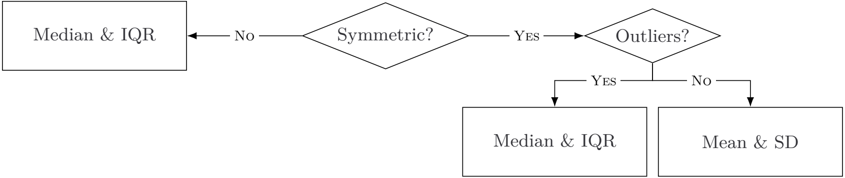
(a) Calculate the mean and median.

(b) What does the relationship between mean and median suggest about the shape?

(c) Calculate the standard deviation.

(d) Which summary statistics are most appropriate?

Decision Guide: Choosing Summary Statistics



Summary: Properties of Statistical Measures

Measures of Centre

Mean ($\bar{x}$)
Balance point
Non-resistant

Median
Middle value
Resistant

Mode
Most frequent
Works with categorical data

Know how find it
for a dataset
or from a histogram/
bar graph

Mean — PAIRS WITH — Std Dev

Median — PAIRS WITH — IQR

Std Dev (s)
Avg distance
Non-resistant

IQR
Middle 50%
Resistant

Measures of Spread

Chapter Summary

Measures of Centre

- **Mean ($\bar{x}$)** — Balance point; not resistant
- **Median** — Middle value; resistant
- **Mode** — Most frequent; works for categorical

Measures of Spread

- **IQR** — $Q_3 - Q_1$; resistant
- **Variance (s^2)** — Avg squared deviation
- **Std Dev (s)** — $\sqrt{s^2}$; same units as data

Graphical Summaries

- **Five-Number Summary:** Min, Q_1, Median, Q_3, Max
- **Standard Boxplot:** Whiskers to min/max
- **Modified Boxplot:** Outliers shown separately
- $1.5\times$ **IQR Rule:** Identifies outliers

Choosing Statistics

- Skewed/Outliers $\Rightarrow$ Median & IQR
- Otherwise $\Rightarrow$ Mean & Std Dev

appears in week 4

Exercise: Mean vs. Median

Data: $\{15, 22, 29, 31, 35, 42, 88\}$

(a) Calculate the mean.

(b) Calculate the median.

(c) Which measure better represents the "typical" value? Why?

Exercise: Outlier Detection

Data:

```
12, 14, 15, 16, 17, 18, 19, 20, 45
```

(a) Find Q_1, Q_3, and IQR.

(b) Calculate the fences.

(c) Are there any outliers? If so, identify them.

Exercise: Interpreting a Boxplot

A company reports employee salaries with this five-number summary:

$$\text{Min} = \$35{,}000 \quad Q_1 = \$48{,}000 \quad \text{Median} = \$55{,}000 \quad Q_3 = \$72{,}000 \quad \text{Max} = \$250{,}000$$

(a) What is the IQR?

(b) Is the distribution symmetric, left-skewed, or right-skewed? Explain.

(c) Should the company report the mean or median salary? Why?